DOCUMENT RESUME

ED 171 790 TM 009 267

AUTHOR Fuentes, Edward J.; Wisenbaker, Joseph M.
TITLE The Use of Teacher Rating of Oral English Proficiency As a Covariate in the Analysis of Reading Scores.
INSTITUTION Research Triangle Inst., Durham, N.C. Center for Educational Research and Evaluation.
PUB DATE Apr 79
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Analysis of Covariance; Elementary Education; Group Tests; Individual Tests; Informal Assessment; Language Development; *Language Proficiency; *Migrant Children; *Oral English; *Reading Achievement; Speech Skills; *Student Evaluation; *Test Validity
IDENTIFIERS Comprehensive Tests of Basic Skills; Elementary Secondary Education Act Title I; Mat Sea Cal Oral Proficiency Tests

ABSTRACT

As part of the evaluation of the Elementary and Secondary Education Act (ESEA) Title I Migrant Program, this study investigated whether teacher judgment of oral language proficiency could be used as a covariate in analyzing the reading test scores of migrant children, so that their possible lack of English proficiency could be taken into account when considering their reading achievement. Teacher ratings of the oral English proficiency of 6,000 second, fourth, and sixth grade migrant children were obtained. Ten percent of these children were also given the MAT-SEA-CAL Oral Proficiency Tests, an individually administered test, which served as a validation of the teacher ratings. Structural equation modeling was used to estimate the relationships between the MAT-SEA-CAL test scores, reading scores on the Comprehensive Tests of Basic Skills, and teacher ratings. The findings indicated that teacher judgments may be used as a measure of migrant children's oral language proficiency and as a covariate in the analysis of reading test scores. However, the study only dealt with the use of teacher ratings for group, not individual, decisions; generalizations made on the basis of these results should consider this distinction. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Session 11.27

ED171790

The Use of Teacher Rating of Oral English Proficiency As a Covariate in the Analysis of Reading Scores

Edward J. Fuentes and Joseph M. Wisenbaker

Center for Educational Research and Evaluation
Research Triangle Institute

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Edward J. Fuentes

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM009 267

Paper presented at the 1979 AERA Annual Meeting, San Francisco, California

ABSTRACT

This study represents one part of a large-scale research effort to design and conduct a national evaluation of the ESEA Title I Migrant Program. The major purpose of this investigation was to determine whether teacher judgment of oral language proficiency could be used as a covariate in analyzing the CTBS reading test scores of migrant children so that their possible lack of English proficiency could be taken into account when considering their reading achievement. Related questions of linguistic development and its relationship to reading achievement were also addressed.

In this study, teacher ratings of the oral English proficiency of 6,000 second, fourth, and sixth grade migrant children were obtained. A ten percent subsample of these children were also given the MAT-SEA-CAL Oral Proficiency Tests, an individually administered language test, which served as a validation of the teacher ratings. Structural equation modeling was used to estimate the relationships between the MAT-SEA-CAL test scores, CTBS reading scores, and teacher ratings.

The findings indicate that teacher judgments may be used as a measure of migrant children's oral language proficiency and as a covariate in the analysis of reading test scores. It was noted, however, that the study only dealt with the use of teaching ratings for group, not individual, decisions and that any generalizations made on the basis of the results should consider this distinction.

INTRODUCTION

This study represents one part of a broader research project presently being carried out by the Research Triangle Institute (RTI). The Office of Evaluation and Dissemination (OED), USOE, has contracted with RTI to conduct a national evaluation of the ESEA Title I Migrant (MTI) program. At the time of this writing, all planning and preliminary stages of the study have been completed and the evaluation is currently entering its final phase, i.e., the analysis and reporting of data that have been collected over more than a year's period.

The individuals that the MTI program serves, by providing both academic and nonacademic aid, are children of migrant agricultural workers and fishers. The majority of these children are of Mexican extraction, followed proportionally by blacks and poor whites. The remainder of the population is divided among native Americans, Orientals, Puerto Ricans, and other miscellaneous groups. There are two categories of eligible children for the MTI program: those children who are active migrants; i.e., they cross school district lines periodically as their parents follow employment, and those children classified as inactive migrants; i.e., they have not moved for at least one calendar year. This inactive group continues to be eligible for MTI services for up to five years after they have "settled out." Most of the monies allotted to the migrant program are expended at the elementary grade level, and it is with this level that this paper is concerned.

As part of the overall evaluation study, suitable measures of basic cognitive skills (reading and mathematics) were to be identified for use as outcome and/or descriptive measures. It was recognized from the outset, however, that any attempt to evaluate migrant education, as in previous evaluations of compensatory programs, would be faced with the inherent problems of measuring gains in achievement for a unique population with instruments developed primarily for representative groups of children. Children involved in MTI programs are generally below grade level in achievement, are from the lowest socioeconomic strata, and suffer a disproportionally high rate of school absenteeism. Many are not proficient in, or have difficulties with, the English language. In addition, unlike other children enrolled in compensatory programs, the migrant child's absence from school is more detrimental to his/her education than the number of lost days would ordinarily

imply. The migrant, by definition, moves with the seasons from locale to locale and with each move is likely to suffer the educational, social, and emotional disruption resulting from a lack of interface of educational programs between states, districts, and schools. The Title I Migrant program was initiated to help overcome these special problems and their effects upon the child's educational achievement.

In this regard, one of RTI's initial tasks was to identify instruments that would best measure and reflect the extent to which the MTI program goal has been attained, given the characteristics of the program recipients. In an attempt to identify suitable measures, various strategies were employed. The current literature was reviewed to establish theoretical grounds on which to base rationale for the measurement of certain cognitive areas related to academic achievement; publications which describe and rate instruments by subject area were used to identify candidate measures; technical manuals of selected tests were referred to as the choices narrowed; and various testing and subject experts were contacted at each stage of the selection process as the work progressed. Finally, a panel composed of measurement experts familiar with the problems of testing minority children was established, and a meeting was held in mid-November 1976 to review progress to date and to make any final recommendations.[1] This procedure resulted in the selection of the Comprehensive Tests of Basic Skills (CTBS) reading and mathematics subtests as the primary measures of academic achievement for second, fourth, and sixth grade migrant children participating in the MTI program (Fuentes and Dunteman, 1976).

RTI's evaluation plan called for two data collection points. The first of these took place in early 1978; the second, between January and April 1979. In the interim, the second, fourth, and sixth grade children selected for study were tracked as they moved with their parents in search of work. Educational

---

[1] The test advisory panel consisted of the following individuals: Dr. John B. Carroll, Director of the Thurstone Psychometric Laboratory, University of North Carolina (absent initially, but later contacted for input); Dr. William E. Coffman, Lindquist Professor of Education and Director of the Iowa Testing Programs, University of Iowa; Dr. Robert Linn, Professor of Educational Psychology, University of Illinois; Dr. Floraline Stevens, Acting Assistant Director, Division of Research and Evaluation, Los Angeles Unified School District; Dr. James Vasquez, Center for Chicano Studies, University of Washington; and Dr. Anthony Conger, Department of Psychological Sciences, Purdue University.

histories, based upon the year's tracking, as well as retrospective information extending to academic year 1976-1977, will be constructed for each of the approximately 6,000 children who were administered the CTBS reading and mathematics subtests at data collection points one and two.

In the collection of all achievement data, out-of-level testing was employed. This testing strategy was based upon the results of a feasibility study conducted earlier by RTI which indicated that the use of test forms one level below the children's school grade placement would minimize any possible floor effects while avoiding any generalized ceiling effect (Fuentes and Dunteman, 1977). In addition, to further explain the test results (and in order that comparisons might be made between groups of children with varying levels of English speaking ability in more purely descriptive analyses), an English language measure was sought that could be used as a covariate when dealing with the children's CTBS reading scores. In this manner, it is hoped that the migrant children's possible lack of English proficiency may be considered in assessing the impact of the MTI program on their reading achievement. Accordingly, a search was carried out to identify a suitable measure of English language proficiency. Prior to the conduct of such a search, however, a set of desirable test characteristics, in terms of applicability to migrant children and specific language areas to be measured, had to be identified in order that an informed choice could be made. This entailed a review of the language test literature.

LANGUAGE AND MEASUREMENT

Although there are many possible definitions of language proficiency, a practical one states that "an individual is proficient in a language if he understands and speaks it with as much ease as an adult member of the speech community" (Horvath, 1975). That is, language proficiency is measured along a continuum ranging from no proficiency to that of an adult speaker of the target language. There are, however, two important characteristics of sample members in the MTI study that had to be kept in mind when applying this definition: (1) the MTI sample is made up of second, fourth, and sixth grade children; and, (2) the majority of the sampled children are members of some disadvantaged minority group. The first of these characteristics has important implications for the use and interpretation of language proficiency

measures applied to young children. Since one cannot expect children to be as proficient as adults, the problem of how the child's language development can be evaluated arises. This problem can be dealt with in two ways: by comparing a child's performance to the typically proficient adult and examining the deviation (i.e., criterion referenced approach), or by comparing the child's proficiency to other children of the same age or grade level (i.e., norm referenced approach). The first way might be difficult to operationalize and interpret, since one would be comparing the deviations of young children to some hypothetical adult model and would expect major deviations. The second way requires the existence or, if necessary, the development of adequate age or grade norms; in particular, the comparison of the average proficiency of various subgroups of migrant children of a given age or grade to other groups of non-migrant or perhaps typical children. The second notable characteristic, as stated, concerned the MTI children's minority status. Minority group members are more apt to use a nonstandard dialect when speaking English than are majority group members. Many language proficiency tests penalize individuals for the use of nonstandard dialects. In this regard, sociolinguists caution that speakers of a nonstandard dialect should not be judged less proficient than speakers of a standard dialect. They point out that language is a social tool and as long as effective communication is the end result of any language, distinctions such as "standard" and "nonstandard" made in particular reference to particular dialects of a language are unimportant (Silverman, et al., 1976). Accordingly, it was decided that the language proficiency measure adopted for use in the MTI evaluation study must reflect this philosophy.

Aside from these sample member characteristics, the linguistic components to be measured, the mode of communication to be tested, and the social domain to be sampled had to be considered. In the linguistic realm there are essentially four components that can be measured with proficiency tests. They are: phonology (sound), grammar (form and structure), semantics (meaning), and vocabulary[2]. It is generally conceded that phonology is the least important

---

[2] Though vocabulary (the total number of words in a particular language) is included as a separate linguistic structure within the domain of language assessment, linguistic theory generally classifies it as a subcomponent of grammar.

of the components listed. Burt and Dulay (1973) point out that no studies have demonstrated that pronounciation difficulties predict problems in acquiring and using the basic structures of a language. As an example, Henry Kissinger, who speaks with a heavy German accent, is certainly fluent in the English language. In the case of migrant children, many undoubtedly speak with a Spanish (or other non-English) accent. Such accents have become a source of group identification and attempts to change or discourage Spanish-accented English may be viewed as an unwelcome imposition of outside standards (Horvath, 1975). Vocabulary tests are also overrated. It is naive to think of language as merely composed of words (Roberts, 1970; Burt and Dulay, 1973). A high score in a vocabulary test may only reflect the relationship between middle or upper class membership and word recognition rather than language proficiency. The remaining two domains (grammar and semantics) are considered more important components of language proficiency.

The measurement of language requires the use of communication skills, including reading, writing, listening and speaking. In a more general sense, reading and listening may be subsumed under the receptive mode, while writing and speaking may be categorized as language production competencies. Whether only one or both of these modes are to be tested has to be considered in choosing a language measure, as do the types of behaviors that would be used to assess a given mode.

The last aspect that must be taken into account when choosing a language proficiency measure is the social domain in which a language is spoken. All individuals use a slightly different "code" or register under different social situations to varying degrees of competence. A highly specialized jargon may be used at the place of employment, while a more informal code is used in the home. The same may be said of second language learners, such as migrant children. They will display varying degrees of English language competence depending upon the social domain that the measure samples; be it the language of the home, school, community or that used while conversing with peers. Which social or linguistic domain to sample, in light of the goals of the MTI program, was also considered in the selection process.

THE SELECTION OF A LANGUAGE TEST

The above considerations, coupled with the constraints inherent in any large-scale study, provided the framework for developing a set of criteria

by which to select an English language proficiency measure for the MTI evaluation study. In the case of the MTI program, however, it was possible to restrict these considerations to a specific testing situation and narrow the choices to a subset of instruments. First of all, the MTI study is primarily concerned with the school domain and, thus, no other social domain need be considered. Moreover, the reading communication skills were to be assessed by the appropriate CTBS subtests and were not to be included in the language measure. Further, as stated previously, tests of phonology and vocabulary are not nearly as informative as tests of grammar and semantics and, therefore, were not considered critical in the choice of a measure of language proficiency.

The measure of English language proficiency should, ideally, sample the school domain in both the productive and receptive modes for at least grammar and semantics. At the same time, the large sample size required for national estimates dictated that a group measure be used in the MTI evaluation in order to keep testing costs within reasonable limits. A measure of oral language production cannot be accomplished by group tests, however, thus candidate tests of English language proficiency were further restricted to language measurement within the receptive mode. Other restrictions affecting test selection relate to acceptable psychometric qualities of the measures (e.g., validity and reliability), adequate standardization procedures, and representative norming samples. With these criteria in mind, a thorough search for appropriate group-administered proficiency tests was undertaken. Sources investigated included publications by the Center for School Evaluation (Hoepfner, et al., 1976), a report issued by the American Institutes for Research dealing with the evaluation of bilingual education (1975), Oral Language Tests for Bilingual Students (Silverman, et al., 1976), Language Proficiency Testing (Horvath, 1975), Non-biased Assessment of Minority Group Children (University of Kentucky, undated), as well as numerous personal contacts with test authors, publishers, and language assessment experts. When these sources proved inadequate or inappropriate, the technical manuals of the tests were consulted. Only ten group-administered tests were identified as possible candidates.

The results of this search indicate that the state of the art for group-administered language proficiency measures is not encouraging. Indeed, a report by the American Institutes for Research (1975) on the selection of a group measure of English language proficiency for elementary school children

states that a thorough evaluation of available instruments failed to identify an appropriate test. The limitations of the available group tests forced the consideration of alternative methods of gathering the language information required for the MTI program evaluation. Because of cost and scheduling constraints noted earlier, these alternatives did not include the use of individually-administered tests for the entire study sample. It was decided instead that, if feasible, teacher ratings of the children's oral English language proficiency would be used in lieu of either group or individual tests. Accordingly, a five-point rating scale ranging from no English proficiency to full proficiency for a child of the given age was developed (see Appendix A). In developing this scale, those linguistic and conceptual areas and sample member considerations used in developing the test selection criteria were again taken into account. All migrant children who were administered the CTBS were also rated on their oral English language proficiency by their teachers or other informed school personnel.

Some validation of the teacher ratings, however, was necessary if they are to serve as the measure of the children's English language proficiency in all future analyses of the MTI evaluation study data. Consequently, a plan was devised in which ten percent of the approximately 6,000 children who had been rated by their teachers and administered the CTBS would also take an appropriate individually-administered language test. The scores on the individually-administered test for the 600 or so children would be compared with the teacher ratings for this same subsample in an attempt to "validate" teacher judgments. Given this research strategy, an effort was launched to identify a suitable individually-administered oral English language test to be used in the teacher ratings validation.

The individual English language measure selected for use in the validation study was chosen through the same criteria developed for the group measures. The only difference was that it was now possible to also include a measure of the productive mode. The same sources used in the selection of group tests were also used to identify appropriate individual instruments against which to apply the selection criteria. As a result, the MAT-SEA-CAL Oral Proficiency Test was chosen. This test is made up of the following three subtests: comprehension, sentence repetition, and structured responses. The MAT-SEA-CAL was administered (by specially trained RTI field staff) to the approximately 600 second, fourth, and sixth grade migrant children in the

teacher ratings validation study sample. These tests were later scored at RTI and a reliability study was carried out on the scores.[3] All three subtests proved to be highly reliable with coefficients of .90, .96, and .89 for, respectively, the comprehension, sentence repetition, and structured responses subtests.

TEACHER RATING VALIDATION AND RELATED ISSUES

If teacher judgment of the migrant children's oral English language proficiency is to serve as a proxy covariate for the more time consuming and expensive MAT-SEA-CAL Oral Proficiency Test in the analysis of reading test scores, then a number of characteristics and relationships must be demonstrated. First, the teacher ratings must display acceptable measurement properties. This issue can be resolved, in this case, by calculating and comparing the means and variances of the ratings of the second, fourth, and sixth grade students and by comparing these indices among groups of children for which a priori estimates of oral English language proficiency can be made. For instance, if the teachers, as instructed, are using as the upper limit of their ratings native English speakers of the appropriate age (i.e., the teachers' point of reference for full fluency is changing as children of progressively higher grades are judged), then the means and variances of the ratings should remain relatively unchanged from grade to grade. Conversely, if children from homes in which English is the dominant language are compared to children from Spanish-dominant homes and children from bilingual homes, a predictable difference in mean ratings between these groups should be apparent.[4] Second, there must exist a positive relationship between the three MAT-SEA-CAL subtests and the teacher judgments. This is simply an issue of concurrent validity. Third, the anticipated positive relationship between the reading test scores and the teacher ratings should be completely explained by the relationships involving these two variables and the oral English language subtests. If this is not the case, then the use of the teacher rating as a

---

[3] A generalizability study was conducted (Cronbach, et al., 1972). This approach appears to be superior to the "classical" method of reliability, since it enables one to better isolate, and thus identify, the various sources of "error."

[4] As part of the biographical data collected prior to the administration of the MAT-SEA-CAL, migrant children were classified by home language.

covariate in later analysis will remove effects associated with the dependent variable as well as the confounding effects associated with oral language proficiency. Simply stated, one must be concerned with the possibility that the teachers rated the children not only on the basis of their oral English proficiency but on their English reading achievement as well. If this is the case, then the CTBS scores will be overadjusted (probably downward) when teacher rating is used as a covariate in the analysis of the migrant children's reading achievement.

The existence of a positive relationship between the oral English subtests and the teacher ratings (i.e., concurrent validity) can be readily determined by testing the correlations between these variables for statistical significance. This is a straightforward approach, justifiable with or without the presence of measurement error in the variables. If, indeed, oral English and reading were measured without error, one could also test for the existence of the second necessary relationship discussed above by simply examining the partial correlation between the reading scores and the teacher ratings while controlling for oral English. However, since measurement errors are known to exist, a more rigorous approach for the resolution of this issue is necessary. The approach chosen in the present instance is structural equation modeling as illustrated in Figure 1, which includes as separate components the children's age within grade, the teacher ratings, the three MAT-SEA-CAL subtests, and the CTBS reading scores. The relationships to be tested among these components of interest are also displayed. The critical relationship is represented by the arrow extending from the CTBS reading scores to the teacher rating component.

In addition to permitting an uncontaminated test of the direct and independent relationship between reading test scores and teacher ratings of oral English proficiency, it was hoped that this model would highlight several issues that may be of interest to linguists and educators who deal with limited-English speaking populations. For example, to some extent, the MAT-SEA-CAL subtests and the CTBS reading test may be conceptualized as representing a testable linguistic developmental sequence. The skills measured by the comprehension subtest of the MAT-SEA-CAL may be viewed as necessary preconditions for the development of skills tapped by the sentence repetition subtest. These, in turn, may be necessary for the development of the skills addressed by the structured response subtest. Additionally, one may hypothesize that in

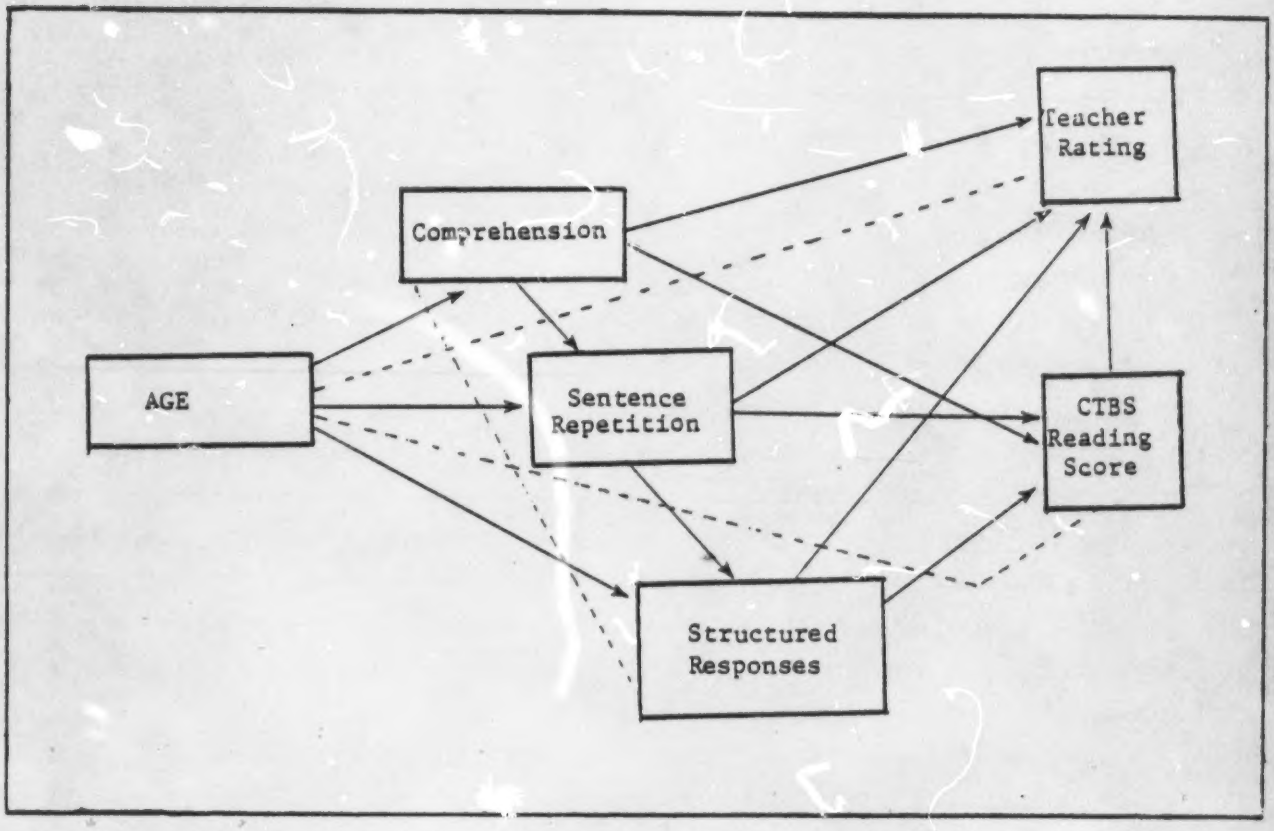

Figure 1. Structural equation model of the hypothesized relationships among the variables of interest.

native English speakers these oral skills naturally precede the reading of English text. By estimating the appropriate structural model it may be possible to determine if such a developmental sequence is evident in the various subgroups of the present sample (i.e., students from English-dominant homes, Spanish-dominant homes, and bilingual homes). For instance, the question of whether second language oral skills are necessary to learn to read in the second language may be explored through the separate analysis of data from native Spanish-speaking students (students from Spanish-dominant homes). A ready contrast is offered by independently estimating the model for native English speakers (students from English-dominant homes). Further, these two models may be contrasted to that developed from data collected on bilingual students (students from bilingual homes).

The results of these analyses may have programmatic significance. If, for example, the interrelationships among the tests of the native Spanish speakers do not support a developmental sequence of English oral skills leading to English reading skills, then it may not be necessary for a program directed toward the development of English reading to incorporate components dealing with oral English. On the other hand, if such a developmental sequence is supported, then it may be said that instruction in oral English appears to be a necessary precondition to instruction in reading English.

ANALYTIC METHOD

In assessing the appropriateness of using teacher rating of oral English proficiency as a covariate in the analyses of CTBS reading scores, the means and variances of each variable involved were calculated within grade level as were simple correlations between each pair of variables and multiple correlation coefficients of MAT-SEA-CAL subtest scores with teacher ratings. These indices were derived to assess both the characteristics of the data produced by the teacher ratings and the relationships between teacher judgment of English proficiency and the MAT-SEA-CAL.

Next, the parameters of structural models for each grade (see Figure 1) were estimated by an approach devised by Jöreskog and van Thillo (1972; Jöreskog, 1973). An advantage of this approach is that it allows the investigator to explicitly incorporate into the model components associated with the reliability of the measures of interest, thereby yielding parameters based upon essentially error-free measures. When multiple measures of each variable exist, the reliability components are routinely estimable as part of the model. In the present case, where only one measure of each variable was available, the reliabilities of the various measures included in the model were obtained from the relevant literature or through reliability studies carried out by RTI. Assuming that these reliability estimates are uniformly valid, the relationships calculated through the structional equations procedure relate to the "true" scores underlying the measures employed, i.e., the error-free measures of the variables. Since the planned analysis of migrant test data will be carried out within grade level, the same strategy was adopted in the present study; hence, the parameters of structural models were estimated, separately, for grades two, four, and six.

In conducting the analyses, an iterative process was employed in determining the final forms of the various structural models. First, the parameters (which may be likened to standardized regression coefficients) of a model were estimated. If, upon inspection, any of the parameter values of this initial model were judged to be approximately zero, a new model was constructed with these same values actually equated or constrained to zero and the remaining non-zero parameters of the resultant simplified model were re-estimated. Next, the parameter representing the true score relationship between the CTBS reading scores and the teacher ratings of oral English proficiency, adjusting for all other variables in the model, was constrained to zero. This, of course, is the condition necessary for the use of the teacher ratings as a covariate in the analysis of the migrant children's CTBS reading scores. In essence, this "ideal" model states that no relationship exists between the CTBS reading score component and the teacher ratings, once all the other variables in the model are accounted for. Finally, a chi-square test of fit was performed in which the observed covariance matrix (the covariances based upon the actual variable scores and ratings) was compared to the expected covariance matrix (the covariances based upon the ideal model parameters). The purpose of this test was to determine whether or not the final model could account for the data. A non-significant chi-square would indicate that such is the case; i.e., that the ideal model is not appreciably different from the initial full model and, therefore, "fits" the data.

Exploration of the issues dealing with linguistic acquisition and its impact on reading involved the separate estimation of non-constrained or full models for subjects from English, Spanish, and bilingual home backgrounds. The relationships among the MAT-SEA-CAL subtests and the CTBS reading score were examined in an effort to identify a distinct language development/reading achievement pattern for each home language group. Here again, analyses were carried out within each of the three grade levels sampled.

RESULTS AND DISCUSSION

All analyses were performed on data collected from a total of 431 migrant children. These individuals, as previously stated, were part of a national sample of approximately 6,000 youngsters participating in the ESEA Title I Migrant program. Although a ten percent subsample of the national sample had

been selected for the present study, a number of individuals were later excluded from the analysis because of missing, incomplete or indeterminate data. Data collection required several days to complete (including make-up examinations), and it was not unusual for some of the youngsters to be absent for part or all of this period or to have changed schools in the interim. Only those children who had taken all reading subtests of the CTBS, had been rated by their teachers on their oral English proficiency and had been administered the MAT-SEA-CAL oral proficiency test were included in the data analysis.

One of the first things to consider, in assessing the advisability of using teacher judgments as a covariate in the analysis of reading scores, is the nature of the teacher rating data. Table 1 presents the means and variances of these ratings by grade and home language group. The number of children that the data are based upon is also presented parenthetically in each cell. Keeping in mind that the teachers were to use typical native English

Table 1

Means ($\bar{X}$) And Variances ($\hat{\sigma}^2$) of Teacher Ratings by Grade And Home Language Group

Grade	Home Language			Total
	English	Bilingual	Spanish	
2	$\bar{X}$ = 4.74 $\hat{\sigma}^2$ = .62 (n = 43)[a/]	$\bar{X}$ = 3.90 $\hat{\sigma}^2$ = .83 (n = 21)	$\bar{X}$ = 3.82 $\hat{\sigma}^2$ = 1.08 (n = 72)	$\bar{X}$ = 4.13 $\hat{\sigma}^2$ = 1.01 (n = 136)
4	$\bar{X}$ = 4.78 $\hat{\sigma}^2$ = .60 (n = 48)	$\bar{X}$ = 4.26 $\hat{\sigma}^2$ = .86 (n = 27)	$\bar{X}$ = 3.98 $\hat{\sigma}^2$ = 1.05 (n = 64)	$\bar{X}$ = 4.30 $\hat{\sigma}^2$ = .94 (n = 139)
6	$\bar{X}$ = 4.79 $\hat{\sigma}^2$ = .61 (n = 39)	$\bar{X}$ = 4.31 $\hat{\sigma}^2$ = 1.00 (n = 29)	$\bar{X}$ = 4.10 $\hat{\sigma}^2$ = 1.04 (n = 88)	$\bar{X}$ = 4.30 $\hat{\sigma}^2$ = 1.00 (n = 156)

[a/] n = sample size

speakers of a given age (without regard to accent) as the benchmarks of full fluency, the means and variances displayed in the "total" column of Table 1 appear to be reasonable. Little difference in these values would be expected over grades if the teachers, as instructed, changed their point of reference for full fluency as migrant children of progessively higher grades were judged. Further, in analyzing the data by home language groups, one would expect children from homes in which English is the dominant language to be rated at the top of the scale and that there would be little change in their means or variances from grade to grade. This is clearly evident from the "English" group column entries. Conversely, one would predict that children from bilingual homes would be rated lower than children from English-dominant homes but higher than children raised in homes where Spanish was the dominant language. This prediction is also supported by the Table 1 data. Finally, as children from bilingual or Spanish-dominant homes receive more exposure to a monolingual English-speaking environment (i.e., as they have more experience in school as indicated by their progression through the grades), their rated oral English should improve. The results shown in Table 1 uphold this hypothesis and, in so doing, provide further evidence that the teacher ratings are operating in a reasonable manner.

On the basis of these analyses, the teacher judgment data are consistent with what may be expected from a sample of migrant children and in this sense teacher ratings are a viable means by which to measure the children's language proficiency. Nonetheless, if teacher ratings are to be used as a surrogate for the MAT-SEA-CAL, it is imperative that a positive relationship also be demonstrated between these ratings and the MAT-SEA-CAL scores. Simple correlations between the MAT-SEA-CAL subtest scores and the teacher ratings were calculated for each grade level and are displayed in Table 2. Additionally, multiple correlation coefficients between the teacher ratings and the MAT-SEA-CAL subtest scores are also reported by grade in Table 3. The data presented in Tables 2 and 3 indicate that the teacher ratings are positively related to the MAT-SEA-CAL scores. Moreover, the coefficients displayed in both tables are moderate to high in value. The only anomalies in the data appear to be the coefficients for the sixth grade. These values are consistently lower than the others, both for the simple and multiple correlations. Nevertheless, even these relationships may be classified as moderate in value; therefore, it may be said that although the relationship between the two measures appears

Table 2

Correlation Coefficients of MAT-SEA-CAL Subtest Scores With Teacher Ratings by Grade

Grade	MAT-SEA-CAL Subtest		
	Comprehension	Sentence Repetition	Structural Responses
2 (n = 136)[a/]	.50	.61	.52
4 (n = 139)	.48	.57	.54
6 (n = 156)	.36	.44	.30

a/ n = sample size

Table 3

Multiple Correlation Coefficients of MAT-SEA-CAL Subtest Scores With Teacher Ratings by Grade

Grade		
2	4	6
.65 (n = 136)[a/]	.64 (n = 139)	.47 (n = 156)

a/ n = sample size

to decrease when estimated on the sixth grade data, a significant positive relationship exists at all grade levels. Given the previous analyses dealing with the characteristics of the teacher rating data, this finding further supports the conclusion that teacher judgments may be used as a substitute for

the MAT-SEA-CAL as a measure of migrant children's oral English language proficiency.

Though it appears that teacher ratings can be used as a measure of migrant children's oral English language proficiency, it still must be demonstrated that the necessary relationships exist among the MAT-SEA-CAL subtests, teacher ratings, and CTBS reading scores (as illustrated in Figure 1) before teacher judgment can be employed as a covariate in the analysis of reading scores. As previously stated, for teacher ratings to be used in this manner, the relationship between the CTBS reading test scores and teacher ratings must be completely explained by the relationships involving these two variables and the MAT-SEA-CAL subtests to avoid removing some effects associated with the dependent variable (reading test scores) when adjusting for the confounding effects associated with the children's oral English proficiency. Structural equation modeling, as described earlier, was the method of choice in addressing this issue.

The final solution of the structural equation modeling analysis is displayed in Tables 4, 5 and 6 for grades two, four, and six, respectively. Each row of a given table may be likened to a separate linear model predicting a single criterion. The entries are standardized regression coefficients estimated from latent variables or "true" score measures. An entire table and its entries represent a matrix of standardized regression values corresponding to the relationships between the latent variables composing a single structural equation model, as illustrated by corresponding Figures 2, 3, and 4. The critical entries are contained in the first row of each table since these values correspond to the variables of major interest (the MAT-SEA-CAL subtests and the CTBS reading scores), which, in turn, predict the teacher ratings. The critical value in this row is that entry which corresponds to the relationship between the CTBS reading scores and the teacher ratings, adjusted for all other variables in the model. If this regression value is constrained to zero and the expected covariance matrix based upon the entire structural equation model does not differ significantly from the observed covariance matrix (the covariances based upon the actual variable scores and ratings), then it may be said that the relationship between the teacher ratings and the CTBS reading scores, when all other variables in the model are accounted for, appears to be zero. This, of course, is the required relationship if teacher ratings are to be used as a covariate in the analysis of migrant students'

reading scores. Each of the models presented in Tables 4, 5, and 6 were subjected to a chi-square goodness of fit test to determine how well they predict the observed data.

Table 4 and Figure 2 present the results of the second grade structural equation modeling analysis and the results of the chi-square test. This model fits the observed data ($\chi^2 = 8.52$ with 6 d.f., $p > .05$) and, therefore, the regression values it includes are considered reasonable estimates of the true score parameters. As can be seen, the solution includes a value of zero for the CTBS/teacher rating relationship when all other variables in the model are accounted for. The use of teacher rating, therefore, as a covariate in the analysis of reading scores is supported by the results of this analysis.

The results of the structural equation modeling analysis for the fourth grade data are displayed in Table 5. These same indices are illustrated in Figure 3 which graphically presents the relationships between the variables. This model also successfully predicts the observed data with the CTBS/teacher rating relationship constrained to zero ($\chi^2 = 2.42$ with 4 d.f., $p > .05$). Thus, again, on the basis of these results, the use of the teacher ratings as a covariate in future reading score analyses appears to be a viable analytic strategy.

Table 6 presents the results of the structural equation modeling analysis based upon the sixth grade data. These results are also displayed graphically in Figure 4. Unlike the solutions presented for the second and fourth grades, the matrix of standardized regression coefficients presented for the sixth grade data contains a non-zero value in the CTBS x teacher ratings cell. Constraining the coefficient representing this relationship to zero resulted in a significant chi-square value of 13.7173 (4 d.f., $p < .05$) for the "ideal" model. The Table 6 results, therefore, do not support the use of teacher rating as a covariate in the analysis of the migrant children's reading test scores.

Although the analysis of the second, fourth, and sixth grade data through structural equation modeling has produced mixed results it may be posited that the relationship between the CTBS reading scores and the teacher ratings, with all other variables in the model accounted for, equals zero. This conclusion, though based primarily on the results of the analyses of the second and fourth grade data, employs input from a variety of sources for support. If, for

Table 4

Standardized Latent Variable Regression Coefficients: Grade Two

Criterion	Predictor				
	Age	Comprehension	Sentence Repetition	Structured Responses	CTBS
Teacher Rating	.0[a/]	.204	.401	.132	.0[a/]
CTBS	.0[a/]	.340	.210	.111	
Structured Responses	.0[a/]	.0[a/]	.702		
Sentence Repetition	-.12	.579			
Comprehension	.0[a/]				

a/ Parameters constrained to zero

$\chi^2 = 8.52$ with 6 d.f., $p > .05$

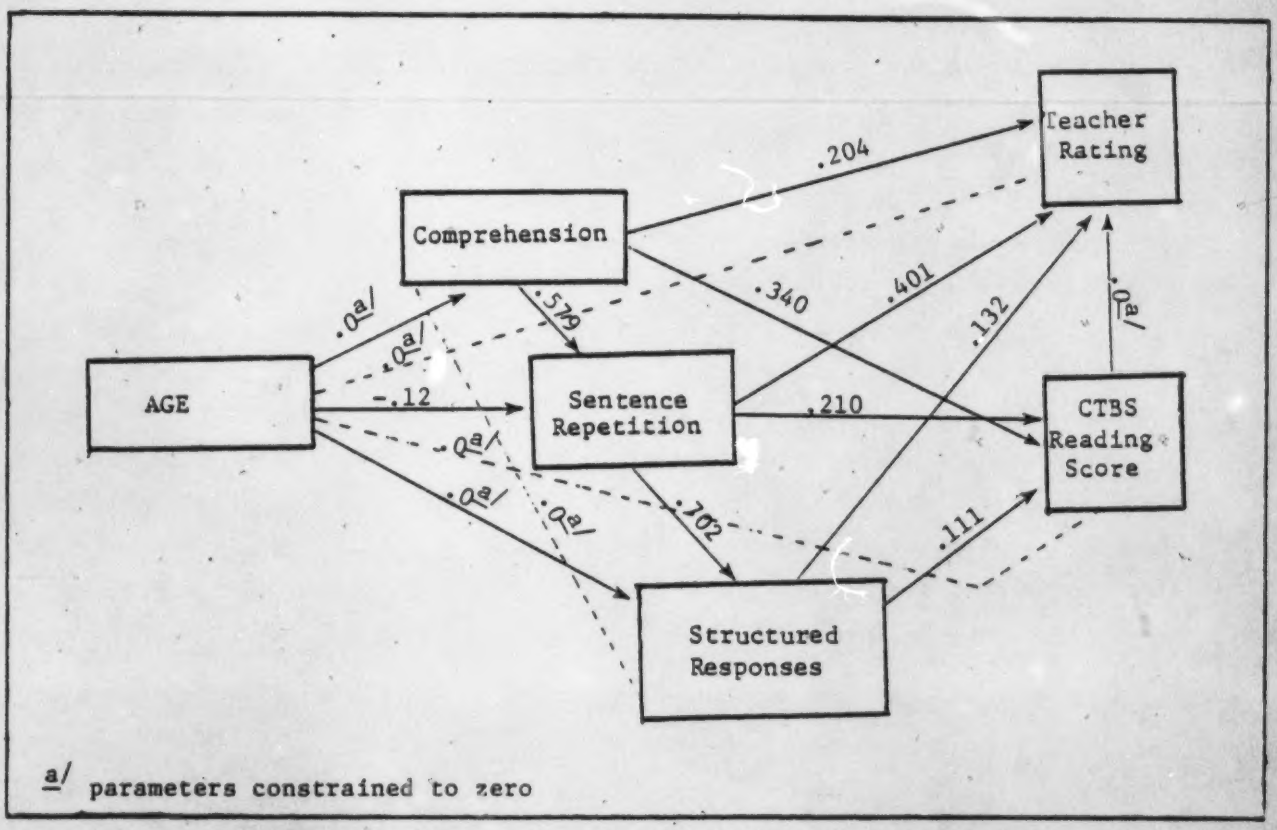

Figure 2. Structural equation model with standardized latent variable regression coefficients: grade two.

Table 5

Standardized Latent Variable Regression Coefficients: Grade Four

Criterion	Predictor				
	Age	Comprehension	Sentence Repetition	Structured Responses	CTBS
Teacher Rating	.0[a/]	.191	.305	.262	.0[a/]
CTBS	.0[a/]	.281	.302	.111	
Structured Responses	.0[a/]	.214	.501		
Sentence Repetition	-.154	.501			
Comprehension	-.231				

a/ Parameters constrained to zero

$\chi^2 = 2.48$ with 4 d.f., $p > .05$

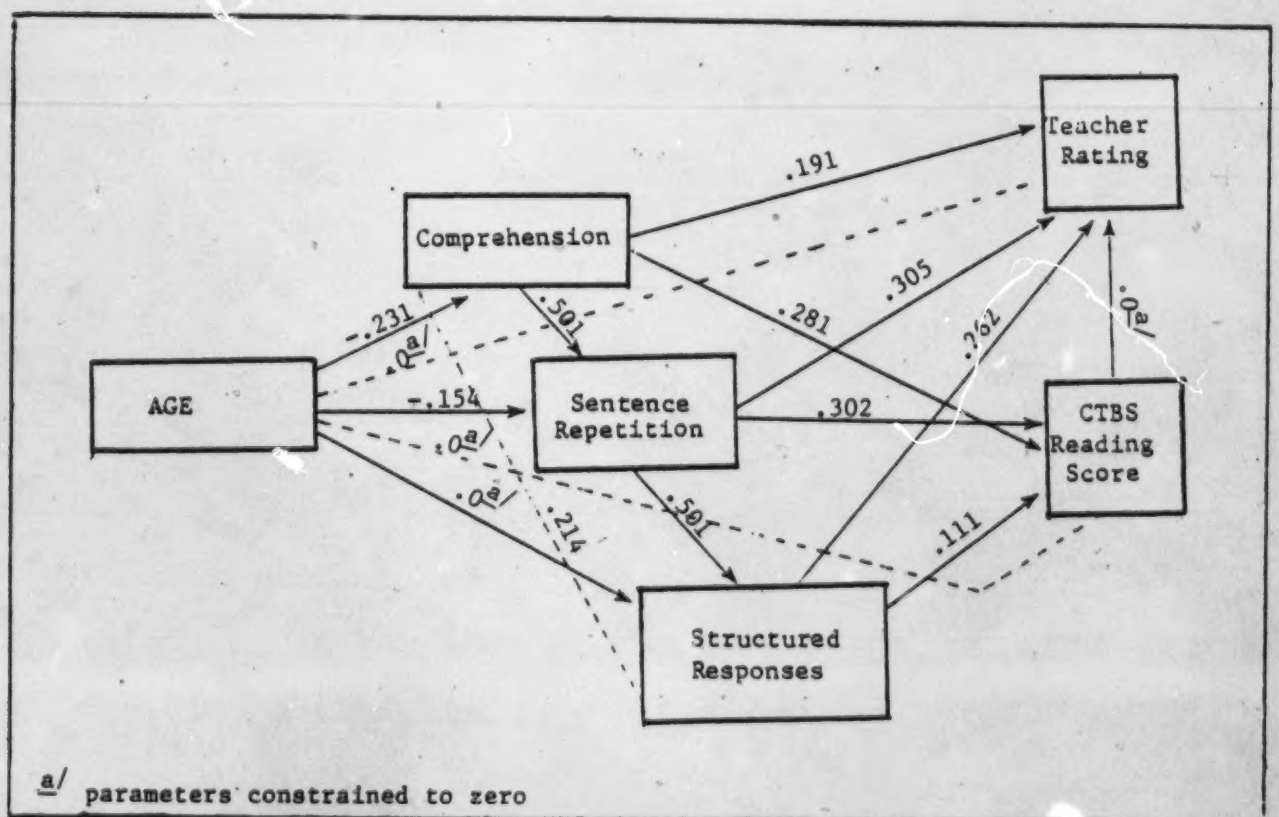

Figure 3. Structural equation model with standardized latent variable regression coefficients: grade four. -19-

Table 6

Standardized Latent Variable Regression Coefficients: Grade Six

Criterion	Predictor				
	Age	Comprehension	Sentence Repetition	Structured Responses	CTBS
Teacher Rating	-.172	.097	.174	.044	.275
CTBS	.0[a/]	.256	.261	.154	
Structured Responses	.0[a/]	.0[a/]	.590		
Sentence Repetition	-.169	.516			
Comprehension	-.215				

a/ Parameters constrained to zero

$\chi^2 = 1.7688$ with 3 d.f., $p > .05$

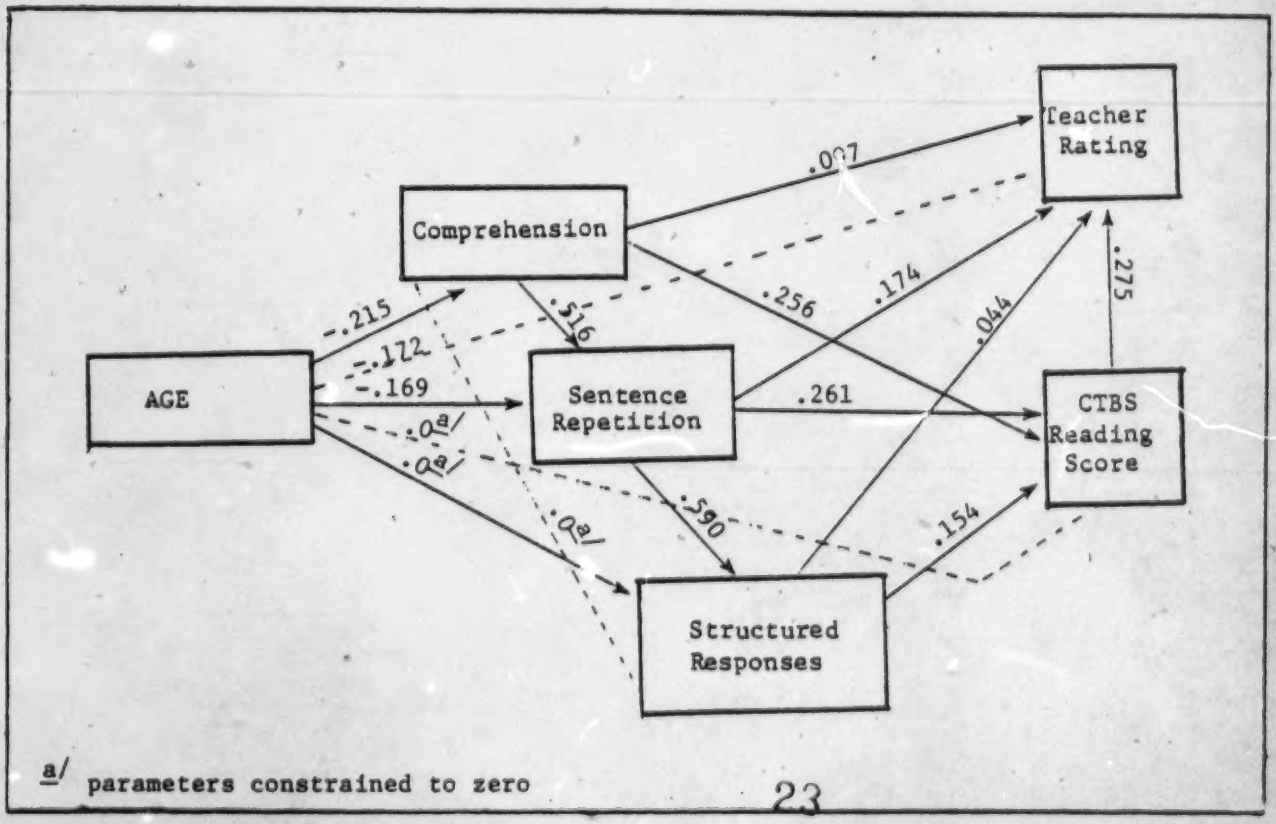

Figure 4. Structural equation model with standardized latent variable regression coefficients: grade six.

example, the first row entries of Table 6 are compared it is apparent that the CTBS x teacher rating coefficient is not only non-zero but is the largest value in the linear model predicting teacher ratings. A non-zero value is, of course, possible but such a relatively large value in the face of the second and fourth grade analytic results prompts a closer examination of other relevant sixth grade data. Table 7 contains the means ($\bar{X}$) and variances ($\hat{\sigma}^2$) of the MAT-SEA-CAL subtest scores for the second, fourth and sixth grade migrant children. Examination of these data across grades reveals that, as grade level increases, there is an increase in the mean and a corresponding decrease in variance. By the sixth grade, the mean values for the comprehension and sentence repetition subtests are approaching their practical upper limits (17 and 39, respectively). This indicates a possible decrease in the reliability of the sixth grade MAT-SEA-CAL scores, which consequently, implies that the MAT-SEA-CAL is no longer adequately measuring the latent trait of oral English ability because of a ceiling effect. The fact that the marginal relationship between the CTBS reading scores and the teacher ratings of oral English is not

Table 7

Means ($\bar{X}$), and Variances ($\hat{\sigma}^2$), of MAT-SEA-CAL Subtests by Grade

MAT-SEA-CAL Subtest	Grade 2 (n = 136)[a/]	Grade 4 (n = 139)	Grade 6 (n = 156)
Comprehension	$\bar{X}$ = 14.40 $\hat{\sigma}^2$ = 5.52	$\bar{X}$ = 15.34 $\hat{\sigma}^2$ = 3.31	$\bar{X}$ = 15.78 $\hat{\sigma}^2$ = 2.62
Sentence Repetition	$\bar{X}$ = 30.54 $\hat{\sigma}^2$ = 65.12	$\bar{X}$ = 34.77 $\hat{\sigma}^2$ = 22.47	$\bar{X}$ = 36.00 $\hat{\sigma}^2$ = 17.39
Structured Responses	$\bar{X}$ = 18.39 $\hat{\sigma}^2$ = 18.66	$\bar{X}$ = 20.63 $\hat{\sigma}^2$ = 14.21	$\bar{X}$ = 20.96 $\hat{\sigma}^2$ = 18.84

[a/] n = sample size

zero, would appear, therefore, to be related more to a change in the psychometric properties of the MAT-SEA-CAL subtests for the sixth grade than to any changes in the criteria upon which the teachers based their judgments. Other supporting evidence for this contention are the previously cited well-behaved changes in the mean teacher ratings from grade two through six, the relatively constant variance in teacher ratings across these same groups, and the reduction in the observed relationship (simple and multiple correlation) between the MAT-SEA-CAL and the teacher ratings for the sixth grade data. The results of all the analyses performed, therefore, strongly suggest that the teacher ratings are a viable measure of the migrant children's oral English proficiency and that their use as a covariate in all future analyses of CTBS reading test scores is appropriate.

The final set of results to be considered in this paper pertain to the more general issues of linguistic development and reading achievement. It had been the authors' intent to demonstrate, through structural equation modeling of the relationships between the MAT-SEA-CAL subtests and the CTBS reading scores for the various home language groups, the link between a child's command of certain language skills and his/her reading achievement. The skills measured by the comprehension subtest of the MAT-SEA-CAL are theoretically necessary preconditions for the development of the skills tapped by the sentence repetition subtest which in turn would seem to be necessary for the development of the skills addressed by the structured response subtest.[5] By checking the relationships between each of these subtests and the CTBS reading scores for different home language groups within each grade, it was hoped that a pattern might emerge which would identify the necessary second-language skills a non-native English speaking child must possess before he/she learns to read. The identification of these language "readiness skills" were to be accomplished by estimating the MAT-SEA-CAL subtest/CTBS relationships through structural equation modeling.

---

[5] The comprehension subtest calls for purely receptive skills since the child is asked to make a choice (by pointing to or marking one of three pictures) in response to a taped stimulus. The sentence repetition and structured responses, on the other hand, call upon the productive mode to varying degrees. In the sentence repetition task the child must repeat a given stimulus; while in the structured response subtest the child is required to complete statements or answer questions designed to elicit certain language constructions.

Table 8

Standardized Latent Variable Regression Coefficients for Grade Two Migrant Children: English Home Language Groups

Criterion	Predictor			
	Age	Comprehension	Sentence Repetition	Structured Responses
CTBS	-.092	.415	.151	.129
Structured Responses	-.014	.211	.457	
Sentence Repetition	-.032	.497		
Comprehension	.329			

Table 9

Standardized Latent Variable Regression Coefficients for Grade Two Migrant Children: Bilingual Home Language Groups

Criterion	Predictor			
	Age	Comprehension	Sentence Repetition	Structured Responses
CTBS	.431	.778	-.209	-.176
Structured Responses	.188	.558	.149	
Sentence Repetition	-.176	.586		
Comprehension	-.117			

Table 10

Standardized Latent Variable Regression Coefficients for Grade Two Migrant Children: Spanish Home Language Groups

Criterion	Predictor			
	Age	Comprehension	Sentence Repetition	Structured Responses
CTBS	-.003	.176	.355	.060
Structured Responses	-.120	.055	.645	
Sentence Repetition	-.143	.495		
Comprehension	-.054			

Table 11

Standardized Latent Variable Regression Coefficients for Grade Four Migrant Children: English Home Language Groups

Criterion	Predictor			
	Age	Comprehension	Sentence Repetition	Structured Responses
CTBS	-.128	.382	.429	.111
Structured Responses	.376	.247	.342	
Sentence Repetition	.183	.305		
Comprehension	.052			

Table 12

Standardized Latent Variable Regression Coefficients for Grade Four Migrant Children: Bilingual Home Language Groups

Criterion	Predictor			
	Age	Comprehension	Sentence Repetition	Structured Responses
CTBS	-.149	.174	.374	.301
Structured Responses	.168	.157	.524	
Sentence Repetition	-.150	.642		
Comprehension	.208			

Table 13

Standardized Latent Variable Regression Coefficients for Grade Four Migrant Children: Spanish Home Language Groups

Criterion	Predictor			
	Age	Comprehension	Sentence Repetition	Structured Responses
CTBS	-.014	.273	.312	.081
Structured Responses	-.070	.136	.541	
Sentence Repetition	-.229	.448		
Comprehension	-.319			

Table 14

Standardized Latent Variable Regression Coefficients for Grade Six Migrant Children: English Home Language Groups

Criterion	Predictor			
	Age	Comprehension	Sentence Repetition	Structured Responses
CTBS	-.315	.352	.340	.349
Structured Responses	-.177	.289	.438	
Sentence Repetition	.155	.280		
Comprehension	.213			

Table 15

Standardized Latent Variable Regression Coefficients for Grade Six Migrant Children: Bilingual Home Language Groups

Criterion	Predictor			
	Age	Comprehension	Sentence Repetition	Structured Responses
CTBS	-.092	.293	.018	.314
Structured Responses	.075	.154	.585	
Sentence Repetition	-.266	.618		
Comprehension	-.251			

Table 16

Standardized Latent Variable Regression Coefficients for Grade Six Migrant Children: Spanish Home Language Groups

Criterion	Predictor			
	Age	Comprehension	Sentence Repetition	Structured Responses
CTBS	-.084	.264	.261	.155
Structured Responses	.142	.104	.682	
Sentence Repetition	-.114	.584		
Comprehension	-.282			

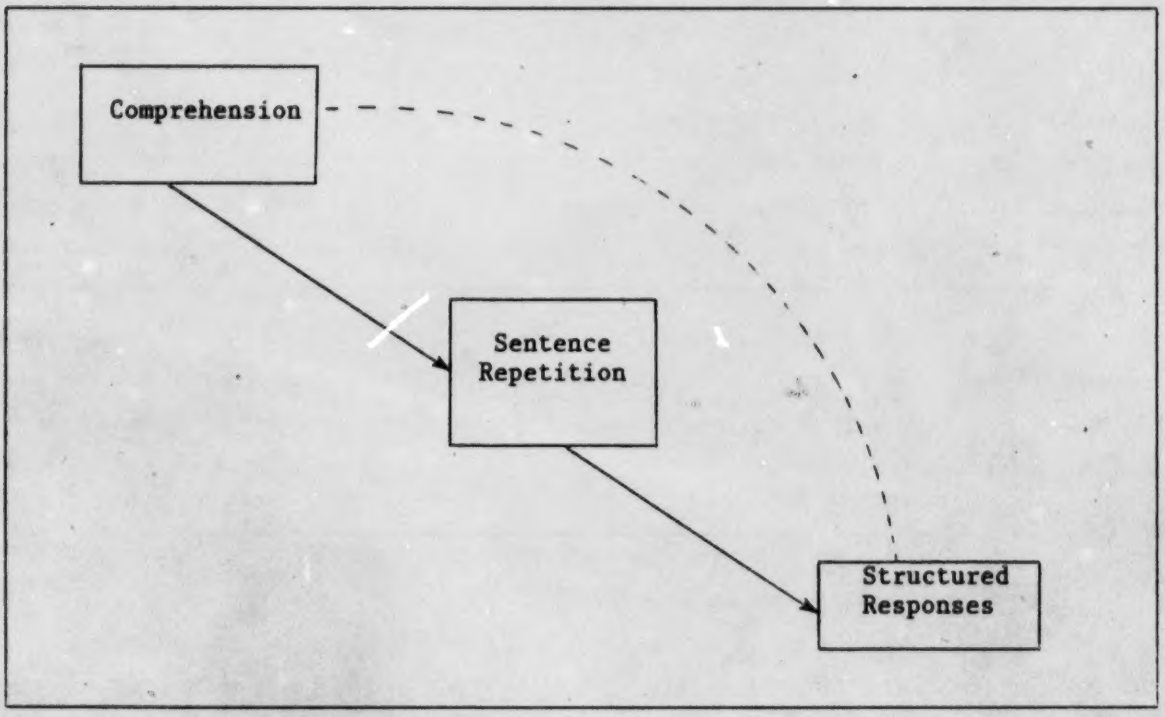

Figure 5. Relationship among the three MAT-SEA-CAL subtests.

Tables 8 through 16 present the results of the structural equation modeling analyses. A separate analysis was conducted for each home language group within grades two, four, and six. The entries are standardized regression coefficients for the latent variables, i.e., they are based upon true score measures of the variables included in the models. An inspection of these tables reveals no compelling patterns in terms of skills (as measured by the MAT-SEA-CAL) for the different home language groups that predict reading achievement (as measured by the CTBS). This is true even when one deals strictly with the absolute magnitude of these values. The issue is obscured further if the magnitude of the differences among coefficients for a given group are taken into account. Many of the coefficients associated with the variables that best predict the children's reading achievement scores do not appear to be significantly different from those coefficients which, in terms of absolute magnitude, are associated with variables of secondary influence. The data collected in this study, therefore, do not lend themselves to the identification of differential linguistic skills development for home language groups as related to reading achievement.

A general pattern among the MAT-SEA-CAL subtests is apparent, however, in the results presented in Tables 8 through 16. This pattern, as illustrated by Figure 5, is consistent from group to group across the three grades. The findings of all structural equation models indicate that the comprehension subtest is the best predictor of the sentence repetition subtest, which, in turn, is the best predictor of structured responses. Stated differently, the contention that the skills measured by the comprehension subtest are requisite preconditions for the development of the skills measured by the sentence repetition subtests, which are necessary for the development of the skills tapped by structured responses has been upheld. This finding speaks well of the underlying theoretical base on which the MAT-SEA-CAL has been built and may have some interest for linguists and educators who deal with limited-English speaking populations.

CONCLUSIONS

This study has addressed several issues and questions of potential importance to individuals concerned with the evaluation of educational programs whose participants include bilingual or limited-English speaking children. The study was conducted on the assumption that any measure used in classifying

children according to their English language proficiency must be carefully chosen on the basis of its technical merits and intended use in future analyses. Moreover, if achievement data are to be adjusted on the basis of the children's language proficiency it must be recognized that a simple relationship between the language competency and achievement may not exist. Other possible underlying structures must be explored in order to have confidence in the results of future achievement data analyses.

It appears, on the basis of an extensive review, that no group-administered language proficiency test meets both the technical and the substantive test selection criteria developed for the migrant program evaluation. In addition, individually-administered instruments are both too expensive and time consuming for use with a large national sample of migrant children. Teacher ratings, on the other hand, proved to be a viable alternative for either group or individually-administered tests. This conclusion, based as it is on the present study, holds if data are to be grouped or aggregated for analysis; however, it does not hold for studies in which the individual is the unit of analysis (i.e., clinical studies concerned with diagnosing language difficulties). If teacher ratings are to be used for individual decisions, then further research must be conducted which addresses this specific issue since the study presented here was not designed to assess the use of teacher ratings for any purpose other than to classify data for further analysis. Based upon the results, such analyses may include the use of teacher judgment of children's oral English language proficiency as a covariate in the analysis of CTBS reading achievement scores.

REFERENCES

American Institutes for Research. Evaluation of bilingual education programs. Palo Alto, California: Addendum No. 1 to considerations and selection of cognitive instruments for the impact study, Author, 1975.

Burt, M. K. and Dulay, H. C. The bilingual syntax measure. Mimeo, 1973.

Cronbach, L. J., Gleser, G. C., Nandar, H., and Rajaratnam, N. The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles. New York: John Wiley and Sons, Inc., 1972.

CTB/McGraw-Hill. Comprehensive Tests of Basic Skills (Form S). Del Monte Research Park, Monterey, California: Author, 1974.

Fuentes, E. J., and Dunteman, G. H. The selection of achievement measures for the evaluation of the Title I Migrant Education Program. Research Triangle Park, N. C.: Research Triangle Institute, October, 1976.

Fuentes, E. J., and Dunteman, G. H. Measuring academic achievement in minority children: the problems of appropriate instrument and test level selection. Paper presented at the Sixteenth Annual Southeastern Conference on Measurement in Education, December 1977.

Fuentes, E. J., and Dunteman, G. H. The selection of English language measures for the Evaluation of the Title I Migrant Education Program. Research Triangle Park, N.C.: Research Triangle Institute, November, 1976.

Hoepfner, R., et al., CSE Elementary School Test Evaluation. Los Angeles: Center for the Study of Evaluation, U.C.L.A. Graduate School of Education, 1976.

Horvath, B. M. Language proficiency testing. Arlington, Va.: Center for Applied Linguistics, 1975.

Jöreskog, K. G. and van Thillo, M. LISERAL: A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables. Princeton, N. J.: Educational Testing Service, December, 1972.

Jöreskog, K. G. A general method for estimating a linear structural equation system. In A. S. Goldberge and O. D. Duncan (eds.). Structural Equation Models in the Social Sciences. New York: Seminar Press, 1973, pps. 85-107.

Matluck, J. H. and Matluck, B. M. MAT-SEA-CAL Oral Proficiency Tests (English). Moscow, Idaho: University of Idaho, School of Education, 1977.

Roberts, E. An evaluation of standardized tests as tools for the measurement of language development. Evanston, Ill.: Northwestern University (mimeo), 1970.

Silverman, R. J., et al. Oral Language Tests for Bilingual Students. Portland, Oregon: Northwest Regional Laboratory, 1973.

University of Kentucky. Non-biased Assessment of Minority Group Children: With Bias Toward None. Lexington, Kentucky: Author, undated.

APPENDIX A

Teacher Rating Scale of Oral English Proficiency

Appendix A

Teacher Rating Scale of Oral Englsih Proficiency

Which of the following statements best describes the child's fluency in oral English?

(Circle one)

Does not speak or understand English	1
Understands some fundamental English. Speaks English only a little if at all	2
Speaks and understands fundamental English sufficiently to participate in elementary conversations. Lack of fluency interferes with his/her classroom work	3
Speaks broken but easily understandable English. Understands most of what is said to him/her in English. Lack of fluency interferes little if at all with his/her classroom work	4
Has reasonable command of the English language for one of his or her age, whether or not he or she speaks with a foreign or regional accent. .	5